Department of Statistics

University of Wisconsin
Madison, Wisconsin

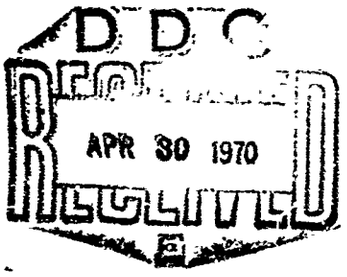
TECHNICAL REPORT NO. 224

February 1970

RESIDUALS AND THEIR VARIANCE PATTERNS

by

Donald W. Behnken*
American Cyanamid Co.
and
Norman R. Draper
University of Wisconsin



Typist: Bernice R. Weitzel

*Statistician-in-Residence at the University of Wisconsin, 1969-70.

1. Summary.

One of the most important aspects of the analysis of data by regression methods is the examination of residuals. This implies the careful inspection of the differences, $e_i = Y_i - \hat{Y}_i$, $i = 1, 2, \ldots, n$, between the observed values y_i, and the corresponding values $\hat{y}_i$ which are predicted by the fitted model at the n observation sites. There are many ways of looking at residuals; see, for example, Draper and Smith (1966, Chapter 3) and Wooding (1969). Important basic techniques are those of plotting the residuals against their corresponding fitted values, or against the corresponding values of the independent variables, or against the corresponding values of "new" variables, and (in all cases) observing the pattern thus formed.

Draper and Smith say that an "ideal" pattern for most plots, which implies no denial of the regression assumptions, occurs when the residuals form a "horizontal band." This is always true for the so called "fixed effect" analysis of variance models. In fitting models with continuous variables, it is usually true within the practical limitations of most plots, but is not precisely true theoretically because the residuals are not independent, nor do they all have the same variance. The purpose of this note is to point out that there is likely to be at least a slight pattern of changing magnitude of the residuals in such plots and that, if such an effect is at all pronounced (as it may well be, given certain properties of the

design matrix) then the variance-covariance structure of the residuals should be taken into account in the analysis.

2. Introduction and Discussion

Suppose the model

$$\underline{y} = \underline{X}\underline{\beta} + \underline{\varepsilon}$$

is fitted by least squares where $\underline{y}$ is an $n\times 1$ vector of observations, $\underline{X}$ an $n\times q$ matrix of known constants, $\underline{\beta}$ a $q\times 1$ vector of unknown parameters, and $\underline{\varepsilon}$ is an $n\times 1$ vector of randomly distributed errors. We make the usual assumptions that $E(\underline{\varepsilon}) = 0$ and $V(\underline{\varepsilon}) = \underline{I}\sigma^2$. The least squares estimate of $\underline{\beta}$ is given by $\underline{b} = (\underline{X}'\underline{X})^{-1}\,\underline{X}'\,\underline{y}$, and the vector of residuals is

$$\underline{e} = \underline{y} - \hat{\underline{y}} = \underline{y} - \underline{X}\underline{b} \qquad (1)$$

$$= (\underline{I} - \underline{R})\,\underline{y} = (\underline{I} - \underline{R})\,\underline{\varepsilon} \qquad (2)$$

where $\underline{R} = \underline{X}(\underline{X}'\underline{X})^{-1}\,\underline{X}'$. Thus the residuals can be regarded as the same linear transformation of the known observations $\underline{y}$ or the unknown errors $\underline{\varepsilon}$. It also follows that

$$E(\underline{e}) = \underline{0} \qquad (3)$$

and

$$V(\underline{e}) = (\underline{I} - \underline{R})\sigma^2 \qquad (4)$$

when the variance assumptions of the model are correct. Since $V(\hat{y}) = \underline{R}\sigma^2$, the variances of the n individual residuals are given by

$$V(e_j) = \sigma^2 - V(\hat{y}_j), \quad j = 1, 2, \ldots, n \qquad (5)$$

$$= (1 - r_{jj})\sigma^2 \qquad (6)$$

where r_{jj} is the jth diagonal element of $\underline{R}$. The pattern of the variances of the residuals is therefore the complement of that of the predicted values $\hat{y}$. It is evident that

$$0 \leq (1 - r_{jj}) < 1 \qquad (7)$$

since $V(e_j)$ is non-negative and r_{jj} is a positive definite quadratic form. $V(e_j)$ is zero only when $e_j = 0$ independent of y, such as in saturated designs (when n=q) or when the peculiarities of the design force $\hat{y}_j$ to equal y_j exactly. For example the residual at the center point of certain second order three-level designs (see Box and Behnken, 1960) with one center point, will always be zero, and V(e) = 0 there.

2.1. Residuals for a first order model with a constant term.

In fitting a straight line model $\eta = \beta_o + \beta x$ we can recall (Draper & Smith, 1966, p.23) that $V(\hat{y})$ increases as the distance of x from the mean value $\bar{x}$ of the observed x's, increases. Figure 1a shows a typical band of 95%

confidence intervals for the expected values of y, derived from five equally spaced observations. As a consequence, residuals at x-sites closer to $\bar{x}$ will have larger variance than residuals further away. Figure 1b shows the pattern of the standard deviation of the residuals for 5 equally spaced observation. Note that we do not extend the "balloon pattern" outside the range of x for which we have observations and, actually, the pattern really consists only of individual verticals intervals at the points at which we have observations. These intervals are drawn in Figure 1b to be of width $2\sigma^{-1}\{V(e_j)\}^{\frac{1}{2}}$ (twice the numbers written on the ordinates of the observation sites) and the end-points are joined by a smooth curve simply to show the variation more clearly. The ratio between the center residual standard deviations and that of an outside point is $2^{1/2} = 1.4$. Such variation in standard deviation would usually not be discernible in a typical residuals plot.

The severity of this "ballooning" of $V(e_j)$ depends on the actual values of the x's used in the regression and may or may not be important in a practical problem. If the variances of the residuals varied a great deal, it would be worthwhile to examine the $e_j/(1 - r_{jj})^{1/2}$, instead of the e_j, in the usual residuals plots, and to use the more correct $e_j/\{s(1 - r_{jj})\}^{1/2}$, instead of e_j/s, as the "normal deviate form" of the residuals. In many cases, as we shall illustrate via examples, this refinement is not needed, but in some cases it may be helpful to avoid possible misinterpretations.

The ballooning of the residual variance at the center of gravity of the data will occur in general whenever first order models with constant terms are fitted, i.e. whenever we fit

$$E(y) = \beta_o + \sum_{i=1}^{k} \beta_i x_i. \qquad (8)$$

We can assume without loss of generality that the x_i's are coded so that $\bar{x}_i = \sum_{j=1}^{n} x_{ij}/n = 0$ for i=1, 2, ..., k. Suppose we write $\underline{X} = (\underline{1}, \underline{D})$ where $\underline{D}$ is the usual design matrix. Then $\underline{R} = \underline{1}\underline{1}'/n + \underline{D}(\underline{D}'\underline{D})^{-1}\underline{D}'$, and it follows that

$$V(\hat{y}_j) = [1/n + t_{jj}]\,\sigma^2 \qquad (9)$$

where t_{jj} is a positive definite quadratic form. Therefore

$$V(e_j) = [(n-1)/n - t_{jj}]\,\sigma^2 \qquad (10)$$

and $t_{jj} = 0$ only if the jth row of $\underline{D}$ is at the centroid where $x_{ij} = 0$ for all i. $V(e_j)$ must increase monotonically away from this minimum since it is a true quadratic in the k independent variables of the first order model. Hence for these models

$$0 \leq V(e_j) \leq [(n-1)/n)]\sigma^2. \qquad (11)$$

2.2. Residuals for first order models without constant terms.

When a first order model without a β_o term is fitted, the ballooning pattern is replaced by a decreasing V(e) as the distance from the actual origin increases. Equations (6) and (7) hold, with the maximum V(e), of value σ^2, being achieved only at the origin. Typically no observations would be taken at the origin in fitting this model, since they do not enter into the values of the estimates at all. If they were taken (e.g. to assist in checking the assumption that $\beta_o = 0$), it is clear that relatively larger residuals should be expected there since $\hat{y}_j$ must always be exactly zero, no matter where the corresponding y_j's lie.

2.3. Residuals for models that are not first order.

When some of the x_i's in a regression model are functions of other x_i's, as, for example, in the second order model,

$$y = \beta_o + \sum_{i=1}^{k} \beta_i x_i + \sum_{i \leq j}^{k} \beta_{ij} x_i x_u + \varepsilon, \qquad (12)$$

equations (6) and (7) hold but no general statements can be made about the location of maxima for V(e) or their number. When a constant term is included, the maximum would hypothetically occur at the point in the k dimensional factor space corresponding to the average of each of the q columns of the $\underline{X}$ matrix. But such a point may not, in fact, exist. For example

in a second order model, when the x_i's are at their average value the x_i^2's are not. The largest residual variance in a second order design may not occur at the center point if the number of replicates there is small enough (see section 3).

2.4. Average value of the variances of the residuals.

Since the average value of the variance of the predicted values at the observation points is given by

$$\overline{V(\hat{y})} = n^{-1} \sum_{j=1}^{n} V(\hat{y}_j) = n^{-1} \operatorname{tr}\{\underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'\sigma^2\} = q\sigma^2/n. \quad (13)$$

where $\underline{X}$ is an $n \times q$ matrix, the average variance of the residuals is

$$\overline{V(\underline{e})} = n^{-1} \sum_{j=1}^{n} V(e_j) = (n-q)\sigma^2/n. \quad (14)$$

Even if the residual variance is reasonably constant, it might still be useful to consider the magnitude of e_i relative to $s/\{(n-q)/n\}^{1/2}$ rather than just s (the estimate of σ) in cases where n is not large relative to q.

3. Examples.

3.1. Straight line in one variable.

In Exercise A, page 35 of Draper and Smith (1966), 11

observations of a response Y occur at 11 equally spaced values of a predictor variable X, at X = -5, -4, ..., 4, 5. The model $y=\beta_o + \beta x + \varepsilon$ is fitted and the variances of the residuals can be shown to be σ^2 times

0.68, 0.76, 0.83, 0.87, 0.90, 0.91, 0.90, 0.87, 0.83, 0.76, 0.68

Thus, the central residual has variance 1.3 times as great as the extreme residuals. (For standard errors, the factor is thus 1.2.) From a practical point of view such variation would not be discernible in an actual residuals plot. The worst variation in the residual standard error one could obtain for the same range with 11 points results from the design with 9 points at the center and one each at ± 5. The equivalent ratios here are 2.2 and 1.5, and a correction might be worthwhile in such a case.

3.2. Straight line through the origin.

Suppose the model $y=\beta x + \varepsilon$ is fitted to observations taken at x=1, 2, 3, 4, 5. The variances of the residuals are σ^2 times

.98, .93, .83, .78, .55

respectively. The steady increase in variance with a factor of 1.8 (1.3 in standard deviation) between lowest and highest might be marginally detectable in a residuals plot and might lead to misleading conclusions unless the possible danger was realized.

3.3. First order rotatable designs.

A common design used for fitting a first order model is a two level factorial or fractional factorial (2^{k-p}) with one or more center points. Since the design is orthogonal and hence first order rotatable, there are but two kinds of sites as far as the variance of a residual is concerned. The variance of b_o is clearly $\frac{1}{n}$, where $n=n_o + 2^{k-p}$, and n_o is the number of replicated center points. Equation (14) can then be used as an easy means of calculating $V(e_f)$, the variance of the residuals at the 2^{k-p} factorial points. We find

$$\frac{1}{n}\{n_o(1 - \frac{1}{n}) + 2^{k-p} V(e_f)\} = \frac{n-(k+1)}{n} \qquad (15)$$

or

$$V(e_f) = \{(n-1)(n-n_o)-nk\}/\{n\, 2^{k-p}\}. \qquad (16)$$

Thus the ratio of the center point residual variance to that at the factorial points is

$$\frac{V(e_{c.p.})}{V(e_f)} = \frac{2^{k-p}(n-1)}{(n-1)(n-n_o)-nk} . \qquad (17)$$

When a 2^3 design with two center points is used, for example, this ratio is 1.7. The standard deviation is therefore 1.3 times as large for the center point residual as for the factorial points and this is of marginal importance.

The simplex design, which is a k+1 point orthogonal

design in k variables, provides perhaps the worst example for a symmetrical design. By the same procedure as used above it is easy to show that

$$\frac{V(e_{c.p.})}{V(e_f)} = \frac{(k+1)(k+n_o)}{n_o} \qquad (18)$$

Here when k=3 and n_o=2 the ratio is 10 and one should not be surprised by very large residuals at the center points. (This is an extreme example, however, because the residuals would also be highly correlated since the residual mean square has only 2 d.f.)

3.4. Second order designs.

3.4.1. Central composite designs.

In fitting a second order model (equation 12) commonly used experimental designs are the central composite designs of Box & Hunter (1957). For a k - factor design these consist of the following points,

x_1	x_2	x_3	...	x_k	
±1	±1	±1	...	±1	2^k or 2^{k-p} factorial points
±α	0	0	...	0	2k axial points
0	±α	0	...	0	
⋮	⋮	⋮	...	⋮	
0	0	0	...	±α	
0	0	0	...	0	n_o center points

(19)

If the value of α is chosen as $2^{(k-p)/4}$, $V(\hat{y})$ is a function only of the distance of the point in the factor space from the center of the design. Designs with this property are called rotatable designs and lead to, at most, three different $V(e_j)$, one for all the factorial points, one for the axial points and one for the center points. In some cases the axial points and factorial points lie on the same hypershere and have the same $V(e_j)$. For the three- and four-factor designs the following results are obtained.

		$\sigma^{-2} V(e_j)$		
Factors (k)	No. of Center Points	Factorial	Axial	Center
3	1	.33	.39	.01
	2	.33	.39	.50
	3	.33	.39	.67
	6	.33	.39	.83
4	1	.42	.42	0
	2	.42	.42	.50
	3	.42	.42	.67
	7	.42	.42	.86

The recommended numbers of center points are 6 (for k=3) and 7 (for k=4). Only with these higher numbers of replicates do the center point residual variances become large relative to the others; the ratios of the standard deviation of a residual at the center to that at a factorial point location are 1.6 (for k=3) and 1.4 (for k=4).

3.4.2. Three level second order designs.

The three level designs of Box and Behnken (1960) have only two different kinds of sites as far as residual variance is concerned, center points and factorial points. Since a

center point residual always has variance $(n_o-1)\sigma^2/n_o$, equation (14) readily provides the factorial point residual variance $V(e_f)$ as

$$\sigma^{-2} V(e_f) = 1 - k(k+3)/(2n_f), \qquad (20)$$

where n_f is the number of factorial points. The design for three factors consisting of the points

x_1	x_2	x_3	
±1	±1	0	12 factorial points
±1	0	±1	
0	±1	±1	
0	0	0	n_o center points

yields the following results:

Factors (k)	No. of Center Points (n_o)	$\sigma^{-2} V(e_j)$ Factorial	Center
3	1	.25	0
	2	.25	.5
	3	.25	.67
	4	.25	.75

The ratio of the standard deviations is 1.6 for $n_o = 3$, and 1.7 for $n_o = 4$.

The three level design for four factors is

x_1	x_2	x_3	x_4	
±1	±1	0	0	} 24 factorial points
0	0	±1	±1	
±1	0	0	±1	
0	±1	±1	0	
±1	0	±1	0	
0	±1	0	±1	
0	0	0	0	n_o center points

This design, however, is a rotation of the four factor central composite design given above in Section 3.4.1 and hence has identical variance values. It is interesting to note that, for all three level designs, $V(e_f)$ is independent of n_o, as equation (20) shows.

The above designs can be regarded as incomplete three level factorial designs. It is also possible to use complete 3^k designs. These designs produce k+1 nominally different kinds of sites; we shall list the $V(e_j)$ for the cases k=2 and k=3 to illustrate the patterns.

Factors	Typical Coordinate	$\sigma^{-2}\, V(e_j)$
2	(±1, ±1)	.19
	(±1, 0)	.44
	(0, 0)	.44
3	(±1, ±1, ±1)	.49
	(±1, ±1, 0)	.66
	(±1, 0, 0)	.74
	(0, 0, 0)	.74

It would appear from these examples that the variation in the standard deviations of residuals from the 3^k levels is not large.

3.5. Some undesigned data examples.

One might expect considerable variation among the $V(e_j)$ when data are taken without following symmetrical experimental designs. Two such examples, taken from Draper and Smith (1966) will be considered next, to provide some insight into what actually does occur.

Example 1. This example (p. 366) uses data previously given by Hald. Here, only the variables x_1 and x_2, which are adequate for representation purposes as described in Draper and Smith (1966, p. 165), are considered. Table 1 provides the x-coordinates, the residuals from a fitted first order model $\hat{y} = 52.577 + 1.468x_1 + 0.662x_2$, the variance of the residuals divided by σ^2, and two columns of standardized residuals. The first of these, e_j/s has, as denominator, the root mean square error $s=(5.79)^{1/2}$, taken from the analysis of variance table; the second is obtained from $e_j/\{\text{Estimated } V(e_j)\}^{1/2} = e_j/s(e_j)$, where the denominator is obtained by substituting $s^2 = 5.79$ for σ^2 in $\{V(e_j)\}^{1/2}$. The pattern of points (x_1, x_2) is shown in Figure 2 with the scale chosen so that the spread of the data points is roughly the same in both x_1 and x_2 directions. We see that the data points are fairly well spread but the point (21,47) is

isolated. Nevertheless this causes no difficulty with this particular set of data and it is clear from Table 1 that the $e_j/s(e_j)$ and e_j/s plots will not be sufficiently different to require use of the former rather than the latter. It is clear however that the relative scaling of the normalized residual for point 10 by a factor of 1.4 compared to that for point 5 could change ones interpretation of a residual plot if large residuals had been obtained at these points.

TABLE 1. Hald data example.

j	x_1	x_2	$\sigma^{-2}\, V(e_j)$	e_j	e_j/s	$e_j/s(e_j)$
1	7	26	.75	-1.6	-.66	-0.77
2	1	29	.73	1.0	.42	0.48
3	11	56	.88	-1.5	-.62	-0.66
4	11	31	.76	-1.7	-.71	-0.81
5	7	52	.92	-1.4	-.58	-0.61
6	11	55	.88	4.0	1.66	1.77
7	3	71	.64	-1.3	-.54	-0.68
8	1	31	.75	-2.1	-.87	-1.00
9	2	54	.82	1.8	.75	0.83
10	21	47	.45	1.4	.58	0.87
11	1	40	.82	3.3	1.37	1.52
12	11	66	.80	0.9	.37	0.42
13	10	68	.79	-2.9	-1.20	-1.36

Example 2. The data for this example are selected from p. 204 of Draper and Smith (1966). The fitted model $\hat{y} = -50.359 + 0.671x_1 + 1.295x_2$ has been used, and the unsymmetrical spread of the design points (x_{1j}, x_{2j}) is shown in Figure 3. The appropriate residuals calculations are shown in Table 2. Note, particularly, that although $\sigma^{-2} V(\hat{y}_j)$ [$= 1 - \sigma^{-2} V(e_j)$] varies a great deal (e.g. 0.06 for point 6 to 0.28 for points 1, 2, and 21), $s^2 V(e_j)$ is quite stable. Although exceptions can exist, it seems reasonable to assume that, as q/n gets small, implying a higher average residual variance, the stability of the residual variances will tend to improve.

Proper standardization of the most suspicious residual, point 21, gives a t-value of -2.73 as compared to -2.32 when s alone is used. The corresponding two tailed percentage points are .014 and .032; although the .032 probability is about halved, the conclusions one would draw are not measurably affected.

TABLE 2. Draper and Smith (1966, p. 205) example.

j	x_1	x_2	$\sigma^{-2} V(e_j)$	e_j	e_j/s	$e_j/s(e_j)$
1	80	27	.72	3.69	1.14	1.34
2	80	27	.72	-1.31	- .40	- .48
3	75	25	.83	4.64	1.43	1.58
4	62	24	.87	5.66	1.75	1.87
5	62	22	.94	-1.75	- .54	- .56
6	62	23	.92	-3.04	- .94	- .98
7	62	24	.87	-3.34	-1.03	-1.14
8	62	24	.87	-2.34	- .72	- .80
9	58	23	.86	-3.36	-1.04	-1.12
10	58	18	.87	2.11	.65	.70
11	58	18	.87	2.11	.65	.70
12	58	17	.80	2.41	.74	.83
13	58	18	.87	- .88	- .27	- .29
14	58	19	.92	-1.18	- .36	- .38
15	50	18	.88	1.48	.46	.49
16	50	18	.88	.48	.15	.16
17	50	19	.88	.19	.06	.06
18	50	19	.83	.19	.06	.06
19	50	20	.84	- .11	- .03	- .04
20	56	20	.94	1.87	.58	.60
21	70	20	.72	-7.53	-2.33	-2.73

3.6. Conclusions.

In making residual plots to check on the validity of model assumptions it appears that, in many situations, li'.tle is lost by failing to take into account the differences in the variances of the residuals. However, it is a potential problem and, since most large regression programs already provide the estimated $V(\hat{y}_j)$, it is a simple matter to add the calculation of the estimated $V(e_j)$ as well and we recommend this as a routine procedure. The normalization of the residuals by their estimated standard deviations can easily be performed as an additional option.

ACKNOWLEDGEMENTS

D.W. Behnken is Statistician-in-Residence at the University of Wisconsin, 1969-70, and is supported by grants from the General Electric Company, and from the Graduate School and the Computing Center of the University of Wisconsin.

N.R. Draper was partially supported by the Office of Naval Research under Contract Nonr-1202(17), Project NR 042-222.

REFERENCES

Box, G.E.P., and Behnken, D. (1960). Some new three level designs for the study of quantitative variables. Technometrics, 2, 455-475. Erratum 3, 576.

Box, G.E.P., and Hunter, J.S. (1957). Multi-factor experimental designs for exploring response surfaces. Ann. Math. Stat., 28, 195-241.

Draper, N.R., and Smith, H. (1966). Applied Regression Analysis. New York: Wiley.

Wooding, W.M. (1969). The computation and use of residuals in the analysis of experimental data. Journal of Quality Technology, 1, 175-188; correction 1, 294.

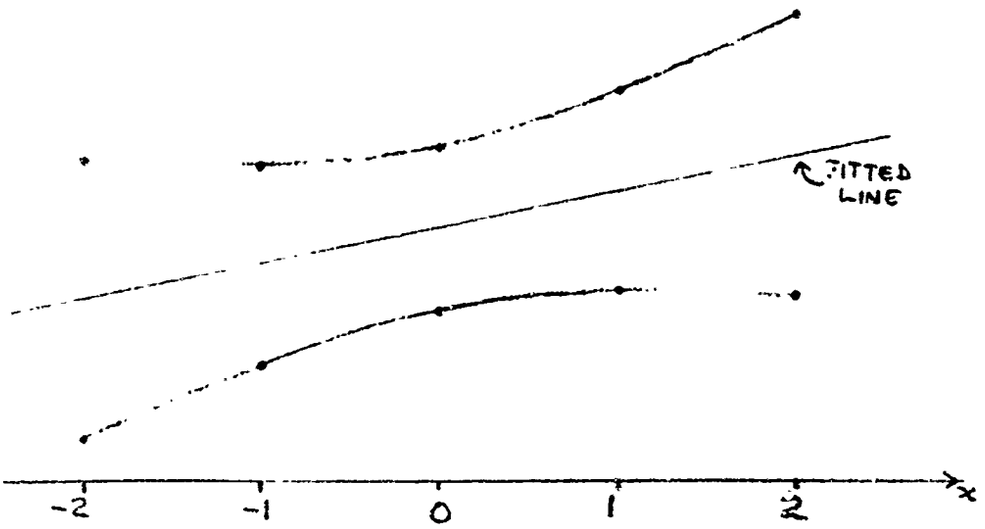


Figure 1a. Typical band of 95% intervals for E(y|x) from five equally spaced observations.

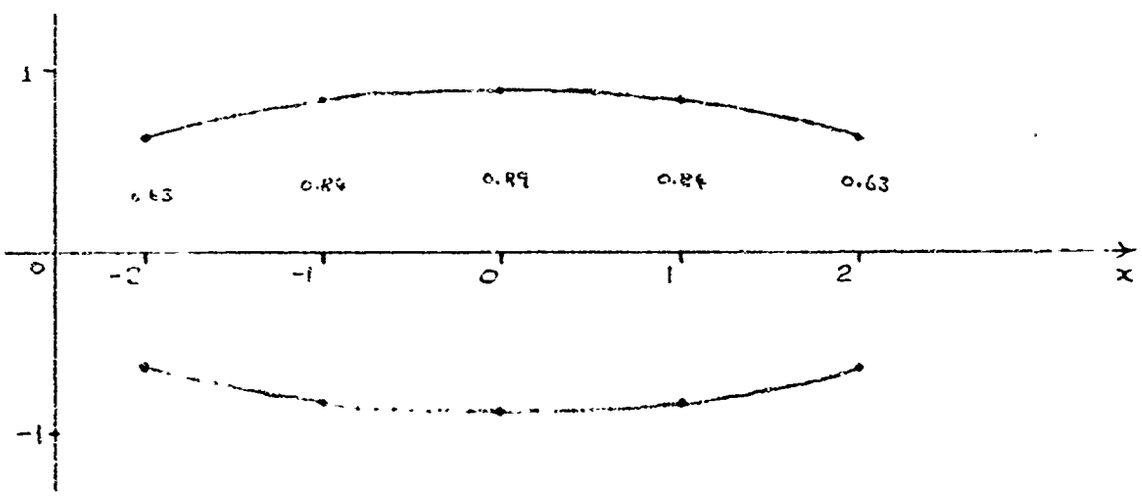


Figure 1b. "Balloon" pattern of standard deviations of residuals from a first order model, for five observations.

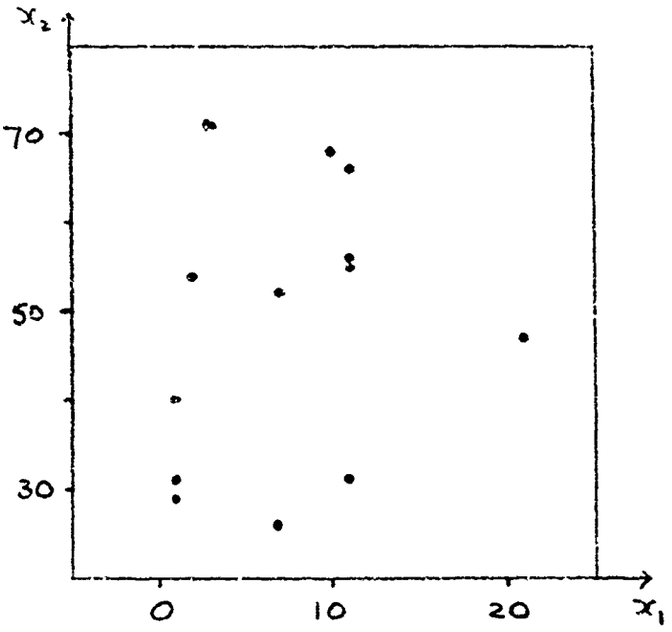


Figure ^. Hald data; (x_1,x_2) scatter plot.

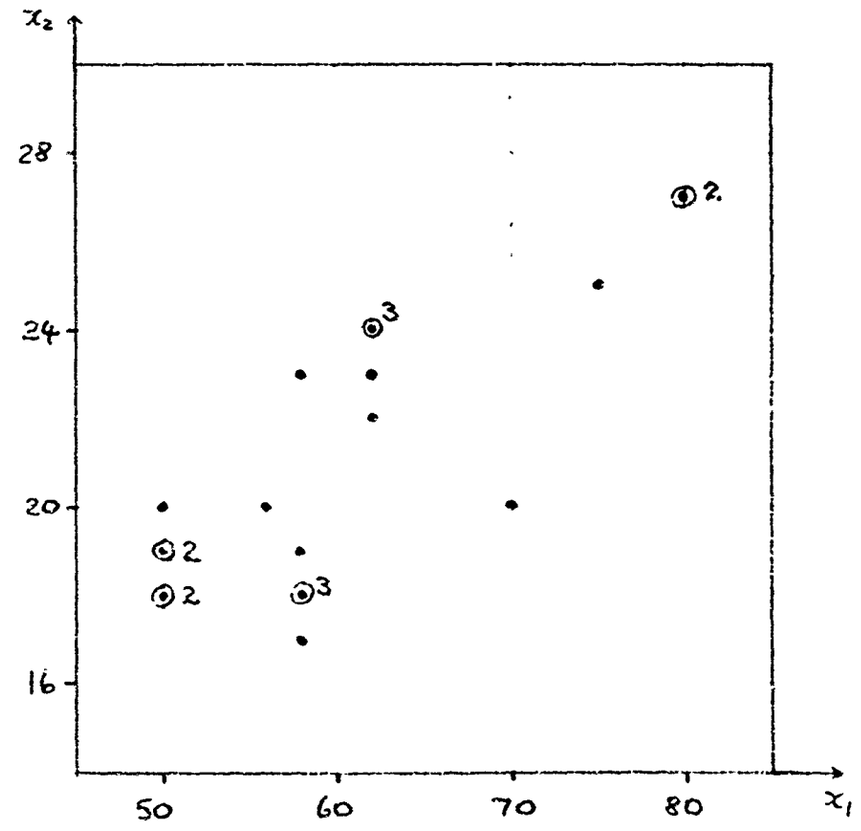


Figure 3. Draper and Smith (1966, p. 205) data; (x_1,x_2) scatter plot. (Some locations have two or three repeat points as indicated.)

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

1. Office of Naval Research U.W. Statistician-in-Residence Program Joint	2a. Unclassified 2b. --

3. RESIDUALS AND THEIR VARIANCE PATTERNS

4. --

5. Donald W. Behnken and Norman R. Draper

6. February, 1970	7a. 22 pp.	7b. 4 references
8a. Nonr-1202	9a. Report No. 224	
8b. NR 042-222	9b. --	

10. Distribution of this document is unlimited.

11. --	12. Office of Naval Research, Washington D.C.

13. Abstract: Residuals are usually examined as though they all had the same variance, and this is not generally true. This paper points out the dangers that can arise from certain types of designs. Concern appears to be usually unwarranted, although there are exceptions.

14. 1. Residuals
2. Regression analysis
3. Variance of residuals

UNCLASSIFIED

Security Classification